

# ExoMol: molecular line lists for exoplanet and other atmospheres

Jonathan Tennyson and Sergei N. Yurchenko

*Department of Physics and Astronomy, University College London, Gower Street, WC1E 6BT London, UK*

Accepted XXXX. Received XXXX; in original form XXXX

## ABSTRACT

The discovery of extrasolar planets is one of the major scientific advances of the last two decades. Hundreds of planets have now been detected and astronomers are beginning to characterise their composition and physical characteristics. To do this requires a huge quantity of spectroscopic data most of which is not available from laboratory studies. The ExoMol project will offer a comprehensive solution to this problem by providing spectroscopic data on all the molecular transitions of importance in the atmospheres of exoplanets. These data will be widely applicable to other problems and will be used for studies on cool stars, brown dwarfs and circumstellar environments. This paper lays out the scientific foundations of this project and reviews previous work in this area.

A mixture of first principles and empirically-tuned quantum mechanical methods will be used to compute comprehensive and very large rotation-vibration and rotation-vibration-electronic (rovibronic) line lists. Methodologies will be developed for treating larger molecules such as methane and nitric acid. ExoMol will rely on these developments and the use of state-of-the-art computing.

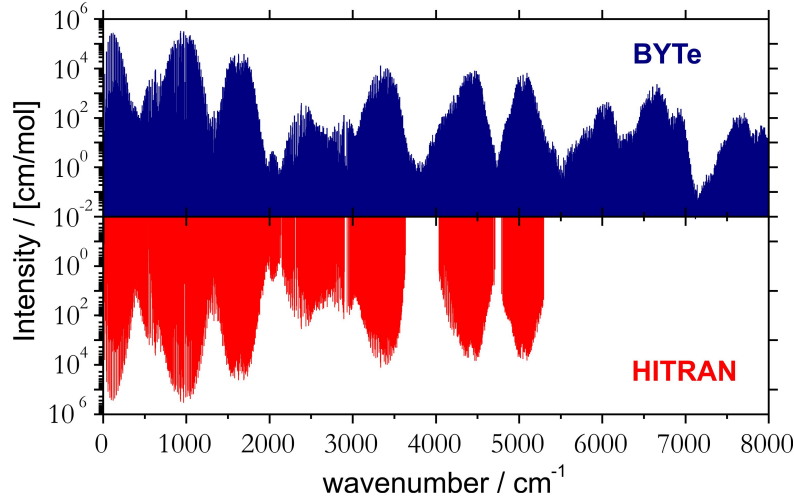
**Key words:** molecular data; opacity; astronomical data bases: miscellaneous; planets and satellites: atmospheres; stars: low-mass

## 1 INTRODUCTION

Most information on the Universe around us has been gained by astronomers studying the spectral signatures of astronomical bodies. Interpreting these spectra requires access to appropriate laboratory spectroscopic data as does the construction of associated radiative transport and atmospheric models. For hot bodies the quantities of atomic and molecular data involved can be very substantial: beyond that which is easily harvested using only laboratory experiments. This problem led, for example, to the establishment of the Opacity Project (The Opacity Project Team & Seaton 1995; The Opacity Project Team & Berrington 1994; Seaton 2005) some 35 years ago with the explicit aim of calculating all the necessary radiative data involving atomic ions which could be of importance for models of (hot) stars. This project was introduced by papers laying the scientific (Seaton 1987) and computational (Berrington et al. 1987) background for the project.

Stars cooler than our own Sun have significant quantities of molecules in their outer atmospheres. These molecules have spectra which are, in general, much richer than those of atoms and atomic ions. They thus both dominate the spectral signature of the cool stars and provide their major opacity sources which, in turn, determine their atmospheric structures. There are even cooler objects which are neither stars nor planets, called Brown Dwarfs. These objects are largely characterised and classified according to the molecular features in their atmospheres. The last decade has also witnessed a rapid escalation in the number of planets orbiting other stars (exoplanets) that have been identified. This number is still increasing rapidly. So far spectroscopic studies of exoplanets are limited in the number, their wavelength coverage and, in particular, their resolution. However it is already apparant from those studies available, which are so far largely confined to hot gas giant planets, that analysing their results will place similar demands on molecular line lists to the requirements of cool stars and brown dwarfs. Modelling and interpreting the spectra of these objects requires data appropriate for temperatures up to about 3000 K.

Since the first detection of sodium in an exoplanet (Charbonneau et al. 2002), exoplanet spectroscopy has made rapid advances. However, even with a rather limited set of molecules detected in exoplanet atmospheres, there remain serious



**Figure 1.** Room temperature ( $T = 296$  K) comparison of laboratory measured spectrum of ammonia, as taken from the HITRAN database, with the line list calculated using program “TROVE” by Yurchenko et al. (2009).

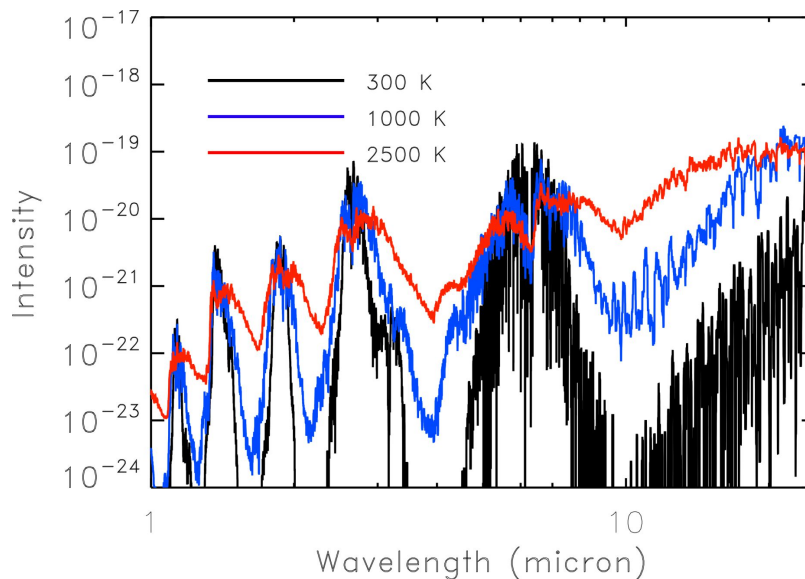
problems with laboratory data. For example, methane was detected in HD189733b by Swain et al. (2008), who lacked the necessary data to determine its quantity; even the presence of methane in other objects remains controversial (Stevenson et al. 2010; Beaulieu et al. 2011).

Planets and cool stars share some common fundamental characteristics: they are faint, their radiation peaks in the infrared and their atmosphere is dominated by strong molecular absorbers. Modelling planetary and stellar atmospheres is difficult as their spectra are extremely rich in structure and their opacity is dominated by molecular absorbers, each with hundreds of thousands to many billions of spectral lines which may be broadened by high-pressure and temperature effects. Despite many attempts and some successes in the synthesis of transition lists for molecular absorbers, reliable opacities for many important species are still lacking.

Determining line lists for hot molecules experimentally is difficult because of (a) the sheer volume of data (maybe billions of lines), (b) the difficulty in obtaining absolute line strengths in many cases, (c) the need to have assigned spectra in order for the correct temperature dependence to be reproduced, (d) the need for completeness, which requires a large range of wavelengths; even at room temperature experimental line lists are often far from complete, see Figure 1 for an example. All this means that a purely empirical strategy is problematic. Instead the plan is to build a reliable theoretical model for each molecule of importance, based on a combination of the best possible *ab initio* quantum mechanical treatment which is then validated by and, in most cases, tuned using experimental data.

Molecular spectra, particularly for polyatomic species, rapidly become extraordinarily rich at elevated temperatures, meaning that the data requirement for a single triatomic molecule can outstrip the entire Opacity Project dataset. Figure 2 illustrates the strong temperature-dependence of the spectrum of water which requires many millions of line to simulate at higher temperatures. Considerable effort has been expended in constructing spectroscopic databases, such as HITRAN (Rothman et al. 2009) and GEISA (Jacquinet-Husson et al. 2011), which provide lists of molecular transitions important at about 296 K. These are appropriate for modelling the atmosphere of our planet and those of the other members of our solar system. However the construction of accurate and complete databases for higher temperatures has been much more partial with most high accuracy studies concentrating on a single species. The present status of this data is reviewed below.

This paper lays out the scientific foundations of a new project, called ExoMol, which aims to systematically provide line lists for molecules of key astronomical importance. These molecules have been selected to be those most likely to be present in the atmospheres of extra-solar planets. In practice they are of importance in many other hot astronomical environments, particularly brown dwarfs and cool stars. The ExoMol project aims to provide a comprehensive database for these objects too. The following section summarises the presently available line lists and illustrates the importance of these line lists by considering some of problems they have been applied to. Section 3 considers the requirements for providing comprehensive data. The molecules concerned are categorised on physical grounds and appropriate methodologies are suggested for each class of problem. Section 4 gives conclusions and perspectives.



**Figure 2.** Absorption spectra of H<sub>2</sub>O given by BT2 (Barber et al. 2007) for  $T = 300, 1000$ , and  $2500$  K.

## 2 THE CURRENT SITUATION

Astronomers interested in molecular line lists use a number of collected data sources. Besides HITRAN and GEISA mentioned above, the JPL (Pickett et al. 1998) and CDMS (Müller et al. 2005) databases provide comprehensive molecular line lists for wavelengths longer than  $30 \mu\text{m}$ . However these databases are aimed at the cool interstellar medium rather than hot sources. HITEMP in both its original (Rothman et al. 1995) and recently updated (Rothman et al. 2010) editions carries data appropriate for modelling molecular spectra at elevated temperatures, but only for five species. Kurucz has extended his well-used atomic opacity tables with data for a number of molecules (Kurucz 2011) but the data for the majority of molecules are approximate and the list of molecules far from complete. Similarly there are partial lists of diatomic opacities provided by the UGAMOP database at the University of Georgia (see [www.physast.uga.edu/ugamop/](http://www.physast.uga.edu/ugamop/)) and the RADEN databank at Moscow State University (Hefferlin & Kuznetsova 1999). The SCAN database also contains line lists for a few diatomics and triatomics (Jørgensen 1996). In summary, while there are a number of sources of molecular line list data, none of them can be considered complete, especially for work at elevated temperatures.

In their review of brown dwarf and very low mass star atmospheres, Allard et al. (1997) found that the majority of molecular opacities available were based on statistical or similarly approximate treatments. Indeed they quote no case where they considered the available molecular line lists to be adequate. This situation has improved somewhat since 1997; Tables 1 and 2 summarise what we believe to be the current situation for line lists of hot diatomic and polyatomic species respectively.

### 2.1 Diatomics

Table 1 lists diatomic line lists which are published, available and fairly complete. Thus, for example, we have omitted the HF line list used by Uttenthaler et al. (2008) as there is no source for this data or, indeed, any details on how it was calculated. Similarly the recent AIO line list of Launila & Berg (2011) contains accurate, measured line frequencies but no transition intensities. In addition, the MARCS model atmosphere code (Gustafsson et al. 2008) contains unpublished molecular line opacities for a number of diatomic species.

It is interesting to consider some of the diatomics that have been treated. Only in a minority of cases, specifically CO, OH and NO which all form part of the HITEMP database (Rothman et al. 2010), have the line lists been constructed essentially on the basis of experimental data, see for example Goorvitch (1994) and Bernath & Colin (2009). A more typical and demanding situation is given by TiO.

TiO is a major opacity source in cool, oxygen-rich stars (Allard et al. 1997). It is an open shell system with several low-lying electronic states which can absorb at near-infrared and red wavelengths, that is close to the radiation peak in a cool star. A number of theoretical studies provided at least partial line lists for this system (Jørgensen 1994; Plez 1998; Alvarez & Plez 1998). At the same time there have been several detailed experimental spectroscopic studies on the system (Gustavsson et al. 1991; Simard & Hackett 1991; Amiot et al. 1995; Kaledin et al. 1995; Ram et al. 1996). Schwenke (1998) combined these studies and data from earlier laboratory spectra (Linton 1974; Hocking et al. 1979; Galehouse et al. 1980; Brandes & Galehouse 1985) with state of the art *ab initio* calculations to give a comprehensive TiO line list containing 37 million lines.

**Table 1.** Recommended available line lists for hot diatomic molecules. Given is the main reference, method: experimental (expt), *ab initio* (ai) or semi-empirical (semi), isotopologues other than the main one, completeness up to a given estimated temperature ( $T^{\max}$ ), number of lines in the line list ( $N$ ) and the electronic source of the data, if available.

Molecule	Ref	Method	Isotopologues	$T^{\max c}$	$N$	Available <sup>d</sup>
C <sub>2</sub>	Kurucz (2011)	semi		5200 K	3459595	Kurucz
CH	Kurucz (2011)	semi		3000 K	71591	Kurucz
CN	Kurucz (2011)	semi		5000 K	1644597	Kurucz
CO	Rothman et al. (2010)	expt		4000 K	113631	HITEMP
CaH	Weck et al. (2003d)	semi		3500 K	89970	UGAMOP
CrH	Burrows et al. (2002)	semi	<sup>50</sup> Cr, <sup>53</sup> Cr, <sup>54</sup> Cr	1300 K	13824	Bernath
FeH	Bernath and co-workers <sup>a</sup>	semi	<sup>54</sup> Fe, <sup>57</sup> Fe, <sup>58</sup> Fe	1600 K	116300	Bernath
HD <sup>+</sup>	Coppola et al. (2011)	ai		all	10120	ExoMol
HeH <sup>+</sup>	Engel et al. (2005)	ai	all <sup>e</sup>	all	8573	ExoMol
LiCl	Weck et al. (2004)	semi		4000K	3357811	UGAMOP
LiH	Coppola et al. (2011)	ai		all	18981	ExoMol
LiH <sup>+</sup>	Coppola et al. (2011)	ai		all	329	ExoMol
MgH	Weck and co-workers <sup>b</sup>	semi		1300 K	23315	UGAMOP
OH	Rothman et al. (2010)	expt		4000 K	41577	HITEMP
NH	Kurucz (2011)	semi		3000 K	36163	Kurucz
NO	Rothman et al. (2010)	expt		4000 K	115610	HITEMP
SiH	Kurucz (2011)	semi		3000 K	78286	Kurucz
SiO	Langhoff & Bauschlicher (1993)	semi				
SiO	Kurucz (2011)	semi		5300 K	1827047	Kurucz
TiH	Burrows et al. (2005)	semi		1800 K	199073	Bernath
TiO	Schwenke (1998)	semi		6200 K	37744499	Kurucz

<sup>a</sup> Dulick et al. (2003); Wende et al. (2010); Hargreaves et al. (2010)

<sup>b</sup> Weck et al. (2003b,a); Skory et al. (2003); Weck et al. (2003c)

<sup>c</sup>  $T^{\max}$  should be considered as a guide indicating the completeness of a line list in question, estimated from the maximal energy  $E^{\max}$  of the line list and the following condition on the Boltzmann factor:  $\exp(-E^{\max}/kT^{\max}) = 5 \times 10^{-7}$ . The latter is an empirical threshold that corresponds to  $T^{\max}$  and  $E^{\max}$  of the BT2 line list (Barber et al. 2006) (see Table 2). 'all' indicates that all bound-bound transitions within the ground electronic state are given.

<sup>d</sup> Data sources:

Bernath: <http://bernath.uwaterloo.ca/XY>, where XY is the chemical formula of the molecule

CDSD databank: <ftp://ftp.iao.ru/pub/CDS-4000>

ExoMol project: [www.exomol.com](http://www.exomol.com)

HITEMP: <http://www.cfa.harvard.edu/hitran/HITEMP.html>

Kurucz CDs, <http://kurucz.harvard.edu/>

UGAMOP project: <http://www.physast.uga.edu/ugamop/>

<sup>e</sup> 'all' means all possible stable isotopologues.

This line list and the corresponding TiO opacity was found to give a very good representation of the TiO absorption in cool stars (Allard et al. 2000) and is now widely used. Schwenke's TiO line list is the largest available for a diatomic by more than an order-of-magnitude.

The next largest available diatomic line list is that for C<sub>2</sub> which is one of a number of diatomics species for which line lists have been provided by Kurucz. Recently, however, there have several new experimental measurements on this system, including the characterisation of entirely new, low-lying electronic bands (Kokkin et al. 2006; Joester et al. 2007; Tanabashi et al. 2007; Nakajima et al. 2009; Bornhauser et al. 2010, 2011). This work has been accompanied by significantly improved *ab initio* electronic structure calculations (Kokkin et al. 2007; Schmidt & Bacskay 2007; Nakajima et al. 2009; Schmidt & Bacskay 2011). Given its importance, C<sub>2</sub> is one of the species we aim to provide an updated line list for.

The important CN radical is represented only by the Kurucz (2011) data in Table 1, which is somewhat approximate. We should note the recent experimental efforts Ram et al. (2010) and by Ram & Bernath (2011) offering accurate but only partial information. More work is therefore needed on this species.

Before turning to larger molecules it is worth considering the FeH and MgH molecules. Both these molecules have been the subject of experimental studies which have provided partial line lists.

In the case of FeH Dulick et al. (2003) present results on rovibronic transitions within the  $F^4\Delta - X^4\Delta$  electronic band. The available line list is based on measured transitions which give the spectroscopic constants of rotational levels belonging to the  $v = 0, 1, 2$  vibrational levels of the FeH  $X$  and  $F$  states; these are then extrapolated to  $v = 3$  and 4, and for  $J (= N + S)$  values up to 50.5, where  $v$  is the vibrational,  $J$  is the total angular momentum,  $N$  is the rotational, and  $S$  is the spin quantum numbers. The line list for this band therefore consists of experimental and extrapolated term values for the 25 vibrational

**Table 2.** Recommended available line lists for hot polyatomic molecules. All line lists are theoretical and designed to be complete up an estimated maximum temperature,  $T^{\max}$ ; the number of lines in millions for the main isotopologue only. For data sources see footnote *d* to Table 1.

Molecule	Ref	Isotopologues	$T^{\max}$	$10^{-6}N$	Available
$\text{H}_3^+$	Neale et al. (1996)	$\text{H}_2\text{D}^+$ (Sochi & Tennyson 2010)	3000 K	12	ExoMol
$\text{H}_2\text{O}$	Barber et al. (2006)	$\text{HDO}$ (Voronin et al. 2010)	3000 K	503	ExoMol
$\text{HCN/HNC}$	Harris et al. (2006)	$\text{H}^{13}\text{CN}$ (Harris et al. 2008)	3000 K	240	ExoMol
$\text{C}_3$	Jørgensen et al. (1989)		3100 K		
$\text{CO}_2$	Tashkun & Perevalov (2011)	all	5000 K	626	CDS
$\text{NH}_3$	Yurchenko et al. (2011a)		1500 K	1014	ExoMol

bands with  $v \leq 4$ . The line list of Dulick et al. (2003) was verified and corrected, by scaling the Einstein  $A$ -coefficients, by Wende et al. (2010) using the high-resolution spectra of red dwarf star GJ 1002. Hargreaves et al. (2010) provide a line list for the  $\text{FeH } E^4\Pi - A^4\Pi$  electronic system near  $1.58 \mu\text{m}$  which combined measured frequencies with *ab initio* calculation of the linestrengths.

$\text{MgH}$ , along with  $\text{CrH}$ , is of potential interest for measuring the presence of deuterium in brown dwarfs (Pavlenko et al. 2008), the so-called deuterium test (Béjar et al. 1999). Extensive experimental studies of  $\text{MgH}$  electronic spectra have been performed by Weck et al. (2003b), Weck et al. (2003a) and Skory et al. (2003). These have been used to compute the complete line list for the  $B' ^2\Sigma^+ - X ^2\Sigma^+$  system of  $^{24}\text{MgH}$ . The list includes transition energies and oscillator strengths over the  $11,850 - 32,130 \text{ cm}^{-1}$  wavenumber range, for all possible allowed transitions from the ground electronic state vibrational levels  $v'' \leq 11$ . This list was computed using the best available *ab initio* potential energies and dipole transition moment function, with the former adjusted to account for experimental data. The status of  $\text{CrH}$  is somewhat similar to this.

It is clear from the above that a mixed experimental and theoretical approach has been the most successful so far. Other experimental datasets are available, for example a very extensive study has recently been completed on  $\text{NiH}$  (Vallon et al. 2009; Ross et al. 2012), and these will provide an appropriate starting point for further line lists.

## 2.2 Polyatomic molecules

Table 2 summarises the available line lists for polyatomic molecules. In the case where several line lists are available, only the recommended one is given.

For polyatomic molecules the main methodology has been theoretical. So far calculations have, of necessity, been performed piecemeal and molecule-by-molecule. For key molecules many line lists may be available; for example, at least seven lists are available for hot water (Allard et al. 1994; Wattson & Rothman 1992; Viti et al. 1997; Partridge & Schwenke 1997; Jørgensen et al. 2001; Schwenke & Partridge 2000; Barber et al. 2006). Studies have shown significant differences between the use of different line lists (see Jones et al. (2002) for example). It is clear that use of complete and spectroscopically accurate line lists is important both for modelling hot astronomical objects and for interpreting their spectra.

The recent CDS-4000  $\text{CO}_2$  line list of Tashkun & Perevalov (2011) extended their earlier work, which was used in the 2010 edition of HITEMP (Rothman et al. 2010), to higher temperatures. This line list was constructed using effective Hamiltonians parameterised using experimental data. Carbon dioxide is a more rigid molecule than the other polyatomics considered in Table 2 and thus a good candidate for treatment using effective Hamiltonians. Very recent, high-temperature, emission experiments by Deprez et al. (2012) suggest that the CDS-4000 line list is indeed the best available at modelling high temperature spectra but there remains some work to be done on this problem. We note that a new theoretical study using methods closer to those advocated here has recently started (Huang et al. 2012).

It is instructive to consider the breadth of applications of the calculated line lists, many of which could not have been anticipated prior to their construction. The polyatomic line lists summarised in Table 2 have all also been used to predict, analyse and assign laboratory spectra of the species, especially at elevated temperatures. These line lists also provide a source of cooling functions (Miller et al. 2010; Coppola et al. 2011) and high-temperature partition functions (Neale & Tennyson 1995; Vidler & Tennyson 2000; Barber et al. 2002) which are important for a variety of astrophysics problems. In addition other applications can be summarised as follows.

$\text{H}_3^+$ . Neale et al. (1996)'s line list and related partition function (Neale & Tennyson 1995) have been used:

- to give all transition intensities for interpreting astronomical observations since there are no laboratory absolute intensity measurements for the spectrum of  $\text{H}_3^+$ ;
- to significantly improve models of cool white dwarf stars (Bergeron et al. 1997);
- to resolve issues with the Jovian energy budget (Miller et al. 2000);

- to probe the role of  $\text{H}_3^+$  in primordial cooling (Glover & Savin 2009);
- to provide stability limits for giant extrasolar planets orbiting near their star (Koskinen et al. 2007);
- to model non-thermal rotational distributions of  $\text{H}_3^+$  both in the interstellar medium (Oka & Epp 2004) and in storage ring experiments (Krechel et al. 2002; Kreckel et al. 2004).

**Water.** The BT2 line list (Barber et al. 2006) has been used:

- to show an imbalance between nuclear spin and rotational temperatures in cometary comae (Dello Russo et al. 2004, 2005) and assign a new set of, as yet unexplained, high energy water emissions in comets (Barber et al. 2009);
- to detect and analyse water spectra in (a) Nova-like object V838 Mon (Banerjee et al. 2005), (b) atmospheres of brown dwarfs (Lyubchik et al. 2007) and (c) FU Orionis objects;
- to calculate the refractive index of humid air in the infrared (Mathar 2007);
- to detect water on transiting extrasolar planets, for which it was completely instrumental (Tinetti et al. 2007) as other available line lists did not contain good enough coverage of the many weak lines that become significant absorbers at high temperatures to make this detection secure;
- for high speed thermometry (Kranendonk et al. 2007), tomographic (Ma et al. 2009) imaging in gas engines and burners and input for models of jet engines (Lindermeir & Beier 2012);
- as input for an improved theory of line-broadening (Bykov et al. 2008);
- to model water spectra in the deep atmosphere of Venus (Bailey 2009);
- to validate the data used in models of the earth's atmosphere and in particular simulating the contribution of weak water transitions to the so-called water continuum (Chesnokova et al. 2009).

**HCN/HNC.** Jørgensen et al. (1985) showed that including HCN opacity in their model of atmospheres of cool carbon-rich stars caused the modelled atmosphere to expand by a factor of 5 and lowered the gas pressure of the surface layers by 1 or 2 orders of magnitude. This finding did much to stimulate detailed work on molecular line opacities. Subsequent line lists (Harris et al. 2002, 2006) have treated the isomerising HCN/HNC as a single species. They have been used

- to detect HNC in the spectra of carbon stars (Harris et al. 2003);
- to constrain C and N abundances in of AGB stars (Matsuura et al. 2005);
- for models of the thermochemistry of HCN (Barber et al. 2002);
- to assign a particularly extensive set of hot, laboratory HCN (Mellau 2011a) and HNC (Mellau 2011b) spectra.

**HeH<sup>+</sup>.** This molecule had been neglected from standard models of helium-rich white dwarfs (which rather surprisingly included both  $\text{H}_2^+$  and  $\text{He}_2^+$ , see for example Stancil (1994)). The line list of Engel et al. (2005) has been used

- for models of the white dwarf stars, particularly helium-rich white dwarfs (Harris et al. 2004);
- to study the effects of early chemistry on the cosmic ray background (Schleicher et al. 2008);
- as a starting point for calculations of end effects in the upcoming KATRIN neutrino mass measurement experiment (Doss et al. 2006).

To add context to the methods discussed below, it should be noted that while the  $\text{H}_3^+$  and  $\text{HeH}^+$  line lists are completely *ab initio*, those for water, HCN/HNC, and  $\text{NH}_3$  used laboratory measurements to improve the procedure. For water and ammonia this came via the use of a spectroscopically determined potential energy surface (PES), while the HCN/HNC calculations replaced the calculated *ab initio* energy levels with observed one where known. Both of these procedures, plus other methods discussed below, take advantage of laboratory high resolution spectra. In contrast, comparisons between dipole transitions intensities computed using completely *ab initio* dipole moment surface (DMS) and benchmark experimental studies have shown that intensities calculated *ab initio* are competitive with, and often more accurate than, laboratory intensity measurements even when they are available (Asvany et al. 2007; Lodi et al. 2011).

### 2.3 Scope of the ExoMol project

A list of species that we plan to consider is given in Table 3. This list is based on current demands for models of exoplanets and brown dwarfs. However it is necessary to be flexible since as the characterisation of exoplanets improves additional molecular line lists are likely to be required.

Broadly speaking, it is possible to separate the spectroscopic demands for hot Jupiters, which have a reducing or hydrogen-rich chemistry, and super-Earth type exoplanets, which can be expected to be oxidising or oxygen-rich. In practice there will be other categories of exoplanets, such as warm Neptunes; however the above division should be sufficient to identify the species for which data are required.

**Table 3.** Exomol list of molecules: Molecular line lists available, fuller lists are given in Tables 1 and 2, and planned categorised by chemistry.

	Primordial (Metal-poor)	Terrestrial Planets (Oxidising)	Giant-Planets & Cool Stars (Reducing atmospheres)
Already available	H <sub>2</sub> , HD <sup>+</sup> , LiH, LiH <sup>+</sup> HeH <sup>+</sup> , H <sub>3</sub> <sup>+</sup> , H <sub>2</sub> D <sup>+</sup>	OH, CO <sub>2</sub> , O <sub>3</sub> , NO H <sub>2</sub> O, NH <sub>3</sub>	H <sub>2</sub> , CN, CH, CO, CO <sub>2</sub> , TiO, YO, HCN/HNC, H <sub>2</sub> O, NH <sub>3</sub>
ExoMol	BeH	O <sub>2</sub> , CH <sub>4</sub> , SO <sub>2</sub> HOOH, H <sub>2</sub> CO, HNO <sub>3</sub>	CH <sub>4</sub> , PH <sub>3</sub> , C <sub>2</sub> , C <sub>3</sub> , HCCH, C <sub>2</sub> H <sub>6</sub> , C <sub>3</sub> H <sub>8</sub> , VO, O <sub>2</sub> , AlO, MgO, BeH, CrH, MgH, FeH, CaH, AlH, SiH, NiH, TiH

Although the atmospheres of other exoplanets are now starting to be probed (Bean et al. 2010; Beaulieu et al. 2011), those exoplanets that are currently the subject of spectroscopic analysis are largely hot Jupiters. For hot Jupiters there is already a major demand for a high quality methane line list (Swain et al. 2008) and there are likely to be demands for line lists of other hydrogenated species such as H<sub>2</sub>S, PH<sub>3</sub>, acetylene, ethane and propane. For super-Earths, that is planets with rocky cores, oxygen bearing species such as ozone and oxides of nitrogen and sulphur are likely to be more important and also will provide bio-signatures (Des Marais et al. 2002). Simulated remote spectra of habitable planets (Kaltenegger et al. 2010) suggest that a variety of molecules including even nitric acid could provide a possible bio-signature. Finally it is already known from models of cool stars that open shell diatomics with low-lying electronic states have clear atmospheric signatures and can play an important role in determining the radiative transport properties of the atmosphere. So far only for TiO (Schwenke 1998) is there a satisfactory line list available.

Table 3 summarises the molecules that are thought likely to be important in the atmospheres of extra-solar planets and cool stars. They are classified according to anticipated atmospheric chemistry, since this is radically different in systems which have no heavy elements (such as bodies formed in the very early universe), are oxygen-rich or are carbon-rich. These species are separated between those for which satisfactory line list, extending to high temperature, are currently available, and those for which line lists are needed. It is clear that, except for primordial chemistries, the to-do list is much the longer. Table 3 has been constructed from the literature (eg Sharp & Burrows (2007); Freedman et al. (2008)) and as a result of extended discussions with several scientists involved directly or indirectly in characterising exoplanets.

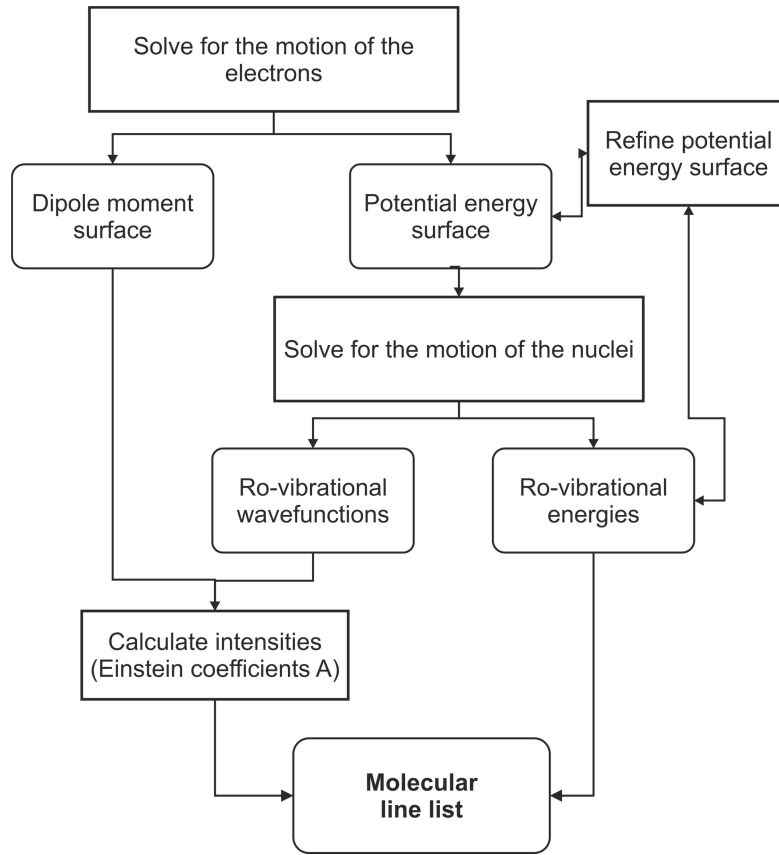
There are significant differences in the physical processes that need to be modelled for different molecules. These are addressed in the next section. Based on the underlying physics that needs to be considered, the problems can be classified as (1) diatomics, (2) triatomics, (3) tetratomics, (4) methane and (5) larger molecules. Special techniques will be required in each case. A final topic will focus on the content, construction and use of the ExoMol database itself.

### 3 METHODOLOGY

Most of the molecules listed in Table 3 are chemically stable and only undergo electronic transitions in the ultraviolet. For these systems it is necessary to consider in detail pure rotational and vibration-rotation transitions within the ground electronic state. However the list also contains a number of open shell diatomic species such as C<sub>2</sub> and FeH. These molecules undergo electronic transitions at near infrared or visible wavelengths; for these systems it is necessary to consider electronic transitions and hence excited electronic states.

Within the Born-Oppenheimer approximation, the calculation of line lists of rotation-vibration transitions for a stable polyatomic molecule can essentially be broken down into the following steps: ground-state electronic structure calculations to give energies and electric dipoles at a series of geometries; interpolation between geometries to create PES and DMS; nuclear motion calculations to provide energy levels and wave functions; calculation of transition dipoles using the wave functions and DMS. Predicted frequencies which arise from such a purely *ab initio* procedure are only accurate enough for present purposes for electronically very simple systems. It is therefore necessary to improve these frequencies using experimental data. This can be done in one of three ways:

- A priori* by tuning the PES by comparison with the results of laboratory high resolution spectroscopic studies. We have developed new procedures for this (Yurchenko et al. 2008, 2011b) which are both efficient and retain the predictive nature of the underlying *ab initio* PES.
- Post hoc* by replacing calculated energy levels with observed ones. As energy levels are not observed directly we will rely on the MARVEL inversion procedure (Furtenbacher et al. 2007) which has been successfully used for water isotopologues (Tennyson et al. 2009, 2010).
- During the calculation: since most of the error is in the vibrational not the rotational energies (Polyansky et al. 1997), using empirical vibrational band origins can significantly improve predicted frequencies for all transitions in the band. The nuclear motion program TROVE (Yurchenko et al. 2007) contains the facility to replace predicted band origins with empirical ones



**Figure 3.** Schematic of the general method that will be used to produce molecular line lists.

part way through the calculation. This facility was used in constructing the recent BYTe line list for  $\text{NH}_3$  (Yurchenko et al. 2011a). In practice all three methods will be used, often in combination. Figure 3 summarises our general methodology.

For each line list the following components are required:

- (i) nuclear motion model, implemented in a computer program, for accurate calculations of rotation-vibration energies and wave functions,
- (ii) accurate PES and DMS,
- (iii) a computational procedure for intensity simulations based on the results of the nuclear motion calculations.

It is generally accepted that *ab initio* DMS computed at high levels of theory provide very reasonable description of the intensities, better for example, than can be obtained by attempting to fit a DMS to measured intensities (Lynas-Gray et al. 1995). Moreover, such properly obtained *ab initio* intensities are, in all but a few cases, superior to data provided by experiment (Lodi et al. 2011).

*Ab initio* PES, however, cannot deliver rotation-vibration energies with sufficiently high accuracy. It is therefore common to empirically refine *ab initio* PESs by least-squares fitting to experimental energies or frequencies to give a “spectroscopic” PES; such potentials can provide theoretical line positions with near-experimental accuracy. When performing such fits it is important to prevent the refined surface from distorting into unrealistic shapes in regions not well characterised by the experimental data. To this end we impose an additional constraint requiring that the refined PES remains relatively close to the underlying *ab initio* PES (Yurchenko et al. 2008). Technically this is done by simultaneous fitting of the potential parameters function both to the experimental (rotation-)vibration energies and to the (lower-weighted) *ab initio* energies (Yurchenko et al. 2003); this is an efficient and, as yet, not widely used procedure.

As mentioned above, the MARVEL (Measured Active Rotation-Vibration Experimental Levels) procedure (Furtenbacher et al. 2007)) provides a rigorous protocol for extracting experimental energy levels from the observed data. So far this protocol has largely been applied to water and the results are only just being incorporated into line lists (Lodi & Tennyson 2012). We will make extensive use of MARVEL during the ExoMol project; note that the “Active” means that the results can be updated as more or improved laboratory measurements become available by simply including these measurements and re-running the process. The original HCN/HNC line list (Harris et al. 2002) was already updated in this fashion (Harris et al. 2006) and extended to  $\text{H}^{13}\text{CN}/\text{HN}^{13}\text{C}$  (Harris et al. 2008) using empirical energy levels, although not ones obtained using MARVEL.



This procedure took advantage of our preferred data structure which separates the line lists into energy and transitions files, and which is discussed in Section 3.6.

There are essentially two different approaches to the construction of the rotation-vibration Hamiltonian. Hamiltonians based on use of an exact kinetic energy (EKE) operator for the nuclear motion of the molecule in question expressed in terms of geometrically defined internal coordinates (Tennyson & Sutcliffe 1982). EKE calculations are potentially very accurate, since the EKE approach can provide an exact (within the Born Oppenheimer approximation) evaluation of the corresponding matrix elements (Tennyson 2011). However, in order to achieve this high accuracy, very substantial computer resources may be needed. Consequently, the EKE approach is the method of choice for triatomics but its use has to be reviewed on a case-by-case basis for larger systems to ensure that the calculations are tractable.

As an alternative, the kinetic energy operator can be defined as a Taylor series in some vibrational coordinates with expansion parameters obtained numerically (Yurchenko et al. 2007). Such a non-EKE approach can be adopted for the larger molecules, some tetratomics and methane. The rotation-vibration coordinates are chosen to minimise the coupling between rotation and vibration. The vibrational coordinates are taken essentially as the Cartesian displacements of the nuclei from their positions when the molecule is at equilibrium. For molecules with large amplitude motions, such as HOOH, we can follow the HBJ method of Hougen et al. (1970) and expand the nuclear kinetic energy operator in terms of the small-amplitude vibrational coordinates from a flexible reference configuration, i.e. around the large-amplitude coordinate. In our extensions of the HBJ theory, the eigenvalues and eigenfunctions of the expanded Hamiltonians are determined variationally by numerical diagonalisation of a matrix representation of the expanded Hamiltonian. This procedure is implemented in the program TROVE of Yurchenko et al. (2007).

For the largest molecules to be considered, those which have five or more atoms (apart from methane), nuclear motion calculations will be performed using new procedures. These will be based on the use of more approximate methods, such as MULTIMODE (Carter & Bowman 1998), DEWE (Mátyus et al. 2007) or an adaption of TROVE, which use normal coordinates and which are more appropriate for fairly rigid molecules. These calculations will not be as accurate as those proposed for smaller systems and therefore extensive tuning to experimental data will be required. Furthermore the potentially huge number of lines required to reproduce a line-by-line spectrum for these species at even moderately high temperature is likely to be prohibitive.

For diatomic and triatomic systems we will generate full line lists in which each rotation-vibration transition is explicitly calculated. For tetratomics and methane, this will be also done where computationally feasible. Otherwise vibrational band intensities and Hönl-London factors will be used to give transition intensities, while the calculated energy levels and hence frequencies will be obtained from calculated, vibrationally averaged rotational constants.

A number of steps will be undertaken to ensure the accuracy of our final line lists. For each system we will initially compute a less comprehensive, low-temperature line list which can be checked for reliability against available laboratory spectra. This step has proved to be fundamental to the success of our recent line list calculations. We note that in the case of methane there are extensive laboratory data, particularly in the near infrared, which have so far defied analysis. It is to be anticipated that our line lists will help resolve these problems as has been the case previously for  $\text{H}_3^+$ , water and the HCN/HNC system.

### 3.1 Diatomics

For many diatomics it is necessary to consider electronic transitions. Best results will rely on the availability of laboratory frequency measurements. Such measurements are, of course, more reliable than we can calculate but are rarely complete, especially for elevated temperatures. Laboratory data are available for the majority of systems such as the extensive dataset of NiH transition frequencies and associated energy levels that have recently become available (Ross et al. 2012). However, it is extremely difficult to construct a complete, high temperature line list only from directly measured data; for example the cited NiH spectra contain no usable information on transition intensities. This makes the construction of a reliable theoretical model essential.

Treatment of the nuclear motion problem for diatomics case is relatively straightforward. For uncoupled electronic states, we will simply use the program LEVEL by Le Roy (2007). In more complicated cases, where strong coupling between different electronic states is important, it will be necessary to consider this coupling explicitly in the calculation. Marian (1995, 2001) has developed a practical theory for such calculations. A program to include couplings between electronic states already exists (Zaharova et al. 2009) but will need to be extensively generalised to cover the many different types of couplings that will be encountered during the project. Given this, the main issue determining the accuracy of the calculations is therefore one of obtaining reliable potential energy curves, curve couplings and transition dipoles.

Our strategy here will be to start from high grade *ab initio* methods: multi-reference configuration interaction (MRCI) expansions based on full-valence complete active space, self-consistent-field reference states utilising large Gaussian basis sets such as aug-cc-pV6Z to resolve valence electron correlation effects. Core and core-valence correlations and scalar relativistic

energy corrections will also be added. Such features are all standard in quantum chemistry programs such as MOLPRO (Werner et al. 2010), Molcas (Aquilante et al. 2010) and Columbus (Lischka et al. 2001).

High quality *ab initio* potential energy and interstate coupling curves will provide the starting point for further refinement. This will take two forms:

- (i) *Ab initio* methods will be used to determine relativistic effects, and in particular relativistic spin-orbit and other couplings between nearby curves.
- (ii) *Spectroscopic* data will be used initially to test curves and then to refine them to give curves that reproduce observed spectra. Much of this can be done with program DPotFit (Le Roy 2006), although this procedure may need extending to deal with molecules whose electronic states are strongly coupled.

Work on diatomic line lists is actively underway and the following paper (Tennyson & Yurchenko 2012) reports line lists for the  $X^2\Sigma^+$  states of BeH, MgH and CaH.

### 3.2 Triatomics

As there are good line lists for the key triatomic species  $\text{H}_2\text{O}$ ,  $\text{CO}_2$ ,  $\text{HCN/HNC}$  and  $\text{H}_3^+$ , relatively few triatomic line lists are planned. Species to be considered include  $\text{H}_2\text{S}$ ,  $\text{C}_3$  and  $\text{SO}_2$ . The nuclear motion problem for these species will be solved using the EKE DVR3D triatomic code (Tennyson et al. 2004), which has already been extensively adapted for the requirement of generating large line lists. For example the algorithm to compute dipole transition intensities was both reworked and parallelised to cope with the requirements of these calculations.

A very accurate spectroscopic  $\text{H}_2\text{S}$  PES was constructed by Tyuterev et al. (2001); this will provide a good starting point although further work will be required to ensure that it remains reliable for higher-lying states. Some work on the role of minor corrections to the PES has already been performed (Barletta et al. 2002). The DMS for  $\text{H}_2\text{S}$  is less straightforward as  $\text{H}_2\text{S}$  has a known feature that the dipole associated with the asymmetric stretch passes through zero close to the equilibrium geometry making the DMS very sensitive to the level of theoretical treatment used to model it. A DMS calculated by Cours et al. (2002) purports to deal with this problem, but in our tests has not been found to be uniformly reliable. Therefore a new higher level theoretical treatment will be needed to give a satisfactory solution to this problem. The procedures developed to produce an essentially exact DMS for water (Lodi et al. 2008, 2011) will be used to determine the level of treatment appropriate for  $\text{H}_2\text{S}$ . In particular, the DMS calculations will use finite differences rather than expectation values which will allow us to test the appropriate level at which to introduce relativistic and other “minor” effects prior to launching a full determination.

A  $\text{C}_3$  line list was produced by Jørgensen et al. (1989), one of the very early ones to be produced. However this line list is no longer accurate by modern standards.  $\text{C}_3$  is a complicated quasi-linear system with an exceptionally flat PES which supports many low-frequency bending modes (Spirko et al. 1997). These have been probed via electronically excited states using stimulated emission pumping (Rohlfing & Goldsmith 1989; Northrup & Sears 1990; Rohlfing & Goldsmith 1990) or laser induced fluorescence (Rohlfing 1989; Baker et al. 1993). Some preliminary work on constructing a new  $\text{C}_3$  line list has been performed (Tennyson et al. 2007) which was based on measurements (Saha & Western 2006) and *ab initio* calculations (Ahmed et al. 2004) performed in Bristol. A first task will be to improve our current PES by performing a new *ab initio* calculation and then tuning to the available data.

The third triatomic for which a line list will be computed is  $\text{SO}_2$ . This is a known constituent of solar system planetary atmospheres (Na et al. 1990).

### 3.3 Tetratomics

Two different codes will be used for treating the tetratomic nuclear motion problem. TROVE (Yurchenko et al. 2007) has already been used to compute an ammonia line list (Yurchenko et al. 2011a) and will be used for phosphine ( $\text{PH}_3$ ) and formaldehyde ( $\text{H}_2\text{CO}$ ). However this code is not appropriate for systems which probe linear geometries such as the linear acetylene ( $\text{HCCH}$ ). For this molecule the EKE code WAVR4 (Kozin et al. 2004) will be employed; indeed acetylene was one of the molecules the code was originally developed to treat (Kozin et al. 2005).

A preliminary acetylene line list was computed by Urru et al. (2010). This calculation gave reasonable results for spectra simulated using vibrational (i.e.  $J = 0$ ) wave functions and vibrational band intensities (Le Sueur et al. 1992) with the rotational fine structure given by vibrational-state dependent rotational constants, and intensities computed using Hönl-London factors. We will aim to produce an improved line list based on a fully-coupled rotation-vibration calculation but first further work will be required on both the PES and DMS. In this context we note the recent emission spectra of hot acetylene obtained by Moudens et al. (2011). In addition we will produce a line list for  $\text{HOOH}$  which is naturally treated using the diatom-diatom coordinate option available in WAVR4.

Phosphine should be amenable to the the same treatment as ammonia and has already been the subject of preliminary, low-temperature studies (Yurchenko et al. 2006; Ovsyannikov et al. 2008; Sousa-Silva et al. 2012). Formaldehyde is the

final tetratomic planned. This has lower symmetry than the other tetratomics discussed so will be computationally the most demanding. However a high quality, spectroscopically-determined PES has recently been developed for this molecule (Yachmenev et al. 2011) which will make an excellent starting point for a line list calculation.

### 3.4 Methane

The detection of methane in exoplanet HD189733b (Swain et al. 2008) was notable for the failure to determine the actual quantity due to the lack of appropriate laboratory data. Similar problems dogged studies of the impact of comet Shoemaker-Levy 9 with Jupiter (Dinelli et al. 1997), and also the interpretation of spectra of brown dwarfs (Homeier et al. 2003), where the desperate resort of modelling them using methane spectra taken from solar system gas giants and Titan has been used (Geballe et al. 1996). Methane is, of course, also an important greenhouse gas as well as being a constituent of many flames. There has been for some time, a clear and pressing need for a comprehensive database of methane transitions. However this is a seriously challenging problem involving the calculation of many billions of vibration-rotation transitions. Advances in computer power and nuclear motion treatments mean that it is becoming technically possible to contemplate a full and systematic computational solution to the methane opacity problem.

There has already been some work in this direction. Schwenke (2002) performed some preliminary studies with a view to computing a line list. More recently Warmbier et al. (2009) did compute a line list using the code MULTIMODE (Carter & Bowman 1998). However this line list has neither the number of transitions nor the accuracy required for models of exoplanet spectra, or indeed the other applications anticipated here. Currently the main source of methane spectra are HITRAN (Rothman et al. 2009), which is only really appropriate for temperatures below 300 K and is then still not complete, or the low-resolution PNNL database (Sharpe et al. 2004). There has recently been significant and coordinated experimental activity to try to understand methane spectra and create corresponding line lists (Albert et al. 2009; Wang et al. 2012), in particular to aid the interpretation of Titan spectra.

Methane is a 10 electron system like water (and ammonia); for water there are well developed procedures for obtaining an ultra-high accuracy PES (Polyansky et al. 2003); application of these procedures should be easier for methane since it is possible to use the faster coupled-clusters approaches, such as CCSD(T), instead of MRCI because of the simpler topology of its PES. Methane has nine degrees of vibrational freedom compared to water which only has three and will therefore require the calculation of significantly more geometries. However methane's symmetry reduces this number by a factor of 24 and use modern computers should allow us to calculate upwards of 50000 points in a few months using MOLPRO (Werner et al. 2010) even with no frozen core, a large (6Z level) or F12 basis set (Hill et al. 2010) while also including allowance for relativistic and adiabatic corrections. This potential can be improved using the extensive experimental datasets referenced above.

The much more difficult steps are the calculation of the 12-dimensional rotation-vibration wave functions and the subsequent calculation of all the associated transition intensities: it is to be anticipated that the final line list will comprise many billions of transitions. There are two possible strategies for solving this problem. In both cases the vibrational calculations will be performed with an upgraded and parallelised version of TROVE (Yurchenko et al. 2007). A strong point of TROVE is the automatic and general treatment of symmetries; this is important not only because maximising the use of symmetry will help to keep the calculation tractable but also because symmetry is necessary to get nuclear spin effects correct when generating spectra.

The comprehensive solution is to use TROVE to simply compute all possible vibration-rotation transitions directly. The calculation of highly rotationally excited states ( $J$  up to 40) and, even more so, the calculation of huge lists of dipole transitions are computationally demanding in the extreme. It is unclear yet, both because the necessary benchmark calculations need to be performed and also because of uncertainty about what computer power will be available, whether this approach will be completely feasible or only so in part.

A more pragmatic, but still reliable, approach is to follow that already used for the preliminary acetylene calculations (Urru et al. 2010). Well-converged wave functions from a  $J = 0$  (i.e. vibration only) calculation will be used to compute (a) vibrational band intensities and (b) vibrational-state dependent rotational constants. The rotational constants will be used to generate the required energy levels and transition frequencies. Vibrational band origins will be combined with so-called Hönl-London factors (well known for the high-symmetry methane molecule) to give the intensity of individual rotation-vibration transitions.

### 3.5 Larger molecules

A characteristic of hot molecules is that their spectra become very congested with many blended lines. For heavier molecules this limit can be reached at room temperature. Thus, for example, the HITRAN database (Rothman et al. 2009) stores all data on species with four or more heavy atoms (about 30 species) as cross-sections rather than fully resolved line lists. However cross-section data are only applicable at the temperature of the measurement, usually room temperature. This is a severe disadvantage which makes the data inflexible. For example, it would be very difficult to use this cross-section data in

atmospheric models of an earth-like planet even when there is only a relatively small temperature differential to earth, say for a super-earth at 350 K.

Conversely the use of variational procedures such as the one outlined above for these heavy systems would be both computationally very expensive and of much lower accuracy than would be required. We therefore propose a rather more empirical approach to address this problem.

The Watson (1968) Hamiltonian provides a general formulation for the nuclear motion of polyatomic systems which is particularly well-suited for use with semi-rigid molecule. This Hamiltonian uses normal modes and, for well behaved systems it is possible to simplify calculations by neglecting high-order coupling terms. This is the basis of the MULTIMODE approach of Carter & Bowman (1998) and DEWE (Mátyus et al. 2007); TROVE can also be adapted to work in this fashion. MULTIMODE has proved capable of giving reasonable results for the rotation-vibration spectra of relatively large molecules Wang et al. (2008).

Calculations based on Watson's Hamiltonian will be used to model room temperature spectra of the species of interest, namely  $\text{HNO}_3$ ,  $\text{C}_2\text{H}_4$ ,  $\text{C}_2\text{H}_6$  and  $\text{C}_3\text{H}_8$ . The calculations will use *ab initio* potentials largely represented by expansions about equilibrium in terms of force constants and higher-order terms. These constants will be calculated *ab initio* as derivatives at equilibrium, empirically-determined if available or a mixture of the two. For some large amplitude modes, such the  $\text{CH}_3$  rotations, full potentials will be used. The PES surface and, if necessary, the associated dipole moments will be systematically tuned to reproduce the measured room temperature cross sections for each system. Initially this will have to be done essentially by trial and error. However with several systems to work on it is anticipated that we should be able to develop systematic procedures for this tuning.

Having developed satisfactory room temperature models for each system, calculations will be repeated at a grid of temperatures to give temperature-dependent cross-sections. Experimental data such as that by Lorono Gonzalez et al. (2010) for  $\text{C}_2\text{H}_4$  will be used to either confirm the model or to provide further input for an improved tuning procedure.

The final output of this model will be cross-sections, since at higher temperatures the number of individual lines will simply be vast and there is little prospect of high resolution spectra of these species being fully resolved in astronomical observations in the near future.

### 3.6 Partition functions

Partition functions are important for models of hot molecules and not altogether straightforward to compute. Extensive compilations of partition functions of astrophysically important species have been made by Irwin (1981) and Sauval & Tatum (1984). These compilations are comprehensive but do not cover all the molecules to considered in the ExoMol project; the partition function values themselves could also, undoubtedly, be improved at higher temperatures. Partition functions which can be considered reliable over an extended temperature range have been constructed for a number of polyatomic molecules including  $\text{H}_3^+$  (Neale & Tennyson 1995), water (Vidler & Tennyson 2000),  $\text{HCN/HNC}$  (Barber et al. 2002) and recently acetylene (Amyay et al. 2011).

In general the partition function is given by

$$Z(T) = \sum_i g_i e^{-c_2 E_i / T}, \quad (1)$$

where  $c_2$  is the so-called second radiation constant and is appropriate when the energy,  $E_i$ , is given as a term value in  $\text{cm}^{-1}$ ,  $T$  is the temperature and  $g_i$  is the statistical weight factor. The statistical weight deserves a special comment.

If the hyperfine structure to be unresolved the statistical weight factor is given by

$$g_i = (2S + 1) g_{\text{ns}}^{(i)} (2N_i + 1), \quad (2)$$

where  $S$  is the total electronic spin angular momentum,  $g_{\text{ns}}^{(i)}$  is the state dependent nuclear statistical weight and  $N_i$  is the rotational angular momentum of the nuclei of the  $i$ th state. If the spin dependent states are resolved the statistical weight factor becomes

$$g_i = g_{\text{ns}}^{(i)} (2J_i + 1), \quad (3)$$

where  $J_i$  is total angular momentum of the  $i$ th state ( $\mathbf{J} = \mathbf{N} + \mathbf{S}$ ). In Eqs. (2) and (3) the hyperfine structure is assumed to be unresolved. We follow the HITRAN convention (Simeckova et al. 2006) and include the entire nuclear statistical weight  $g_{\text{ns}}^{(i)}$  of the molecule explicitly in  $g_i$  and hence the partition function  $Z(T)$ . For example, for BeH where  $^9\text{Be}$  and H have nuclear spins 3/2 and 1/2, respectively,  $g_{\text{ns}}^{(i)} = 8$ . If, as can be assumed for BeH (Tennyson & Yurchenko 2012), the spin-rotation coupling for the ground electronic state  $X^2\Sigma^+$  is unresolved, then each rovibronic states  $i$  is assumed to be doubly  $(2S + 1)$  degenerate. According to Eq. (2) the statistical weight factor of BeH is then given by  $g_i = 2 \times 8 \times (2N_i + 1) = 16 (2N_i + 1)$ . If the spin dependent states are resolved then according to Eq. (3)  $g_i = 8 (2J_i + 1)$ . For the  $^{12}\text{C}^{12}\text{C}$  molecule whose nuclear spin is 0, the statistical weights  $g_{\text{ns}}$  of the symmetric  $s$  and antisymmetric  $a$  rotational levels are 1 and 0, respectively. In the

case of the  $X^1\Sigma^+$  ground electronic state of this molecule  $g_e = 1$  and the statistical weight factors  $g_i$  are  $(2J_i + 1)$  and 0 for the  $s$  and  $a$  states, respectively.

We note that the partition function,  $Z$  can be used estimate the completeness of a line list as a function of temperature. This can be done by comparing the ratio of the partition function computed by summing over all lower-state energy levels used to compute the line list to the accurate partition function (Neale et al. 1996; Barber et al. 2006). Indeed this ratio can even be used make corrections for the missing contribution to, for example, the cooling function (Neale et al. 1996).

### 3.7 The ExoMol database

The backbone of the ExoMol database will be line lists of transitions. However the database will include a variety of associated and ancillary data. These will include energy levels, partition functions, cooling functions and cross sections.

The large amount of data produced in the above calculations, both completed and anticipated, requires the development of strategies for data handling and distribution. Small line lists, which is likely include most diatomics, a simple line list will be stored. However for larger line lists we will employ a data structure which involves organising our final line list into two files: an energy file and a transitions file. The energy file will contain energy levels for each state combined with a number for its position in the file and quantum number assignments, both rigorous and approximate. The transitions file will be arranged in ascending frequencies and will list only the number of the upper and lower state for each transition plus the associated Einstein-A coefficient. This provides a very compact means of representing the data which is essential for efficient use of storage. However, it is likely that the transition file will still need to be split into frequency bins for ease of distribution.

This data structure has other important advantages. It will also allow us to actively update the energy file with measured rotation-vibration energy, or indeed improved approximate quantum numbers, ensuring the best possible for each transition. Indeed Harris et al. (2008) turned an  $\text{H}^{12}\text{CN}/\text{HN}^{12}\text{C}$  line list into one for  $\text{H}^{13}\text{CN}/\text{HN}^{13}\text{C}$  by replacing the energy file with one appropriate for the  $^{13}\text{C}$ ; this approach relied on the not unreasonable assumption that the Einstein-A coefficients do not change significantly between the two isotopologues.

The sheer volume of data contained in the line lists makes them fairly tricky to use. In practice most codes will use some sort of opacity or importance sampling technique to identify key transitions and to discard the rest. We have, however, constructed a set of zero-pressure, temperature dependent cross sections for the key species studied so far (Hill et al. ???). The purpose of these is not to replace the underlying line lists, whose use will remain necessary for detailed studies and analysis of high resolution spectra, but to allow the effects of adding a species to a model to be quickly and efficiently tested. The use of these cross sections avoid the issues of handling huge line lists at the price of assuming local thermodynamic equilibrium (LTE) and some loss of flexibility. The issue turning the line lists into cross sections will be discussed elsewhere (Hill et al. ???).

The results of the calculations outlined above will be a comprehensive database of molecular transitions. The ExoMol database, see [www.exomol.com](http://www.exomol.com), will include not only the line lists, cross sections cooling functions, partition functions and other data generated during the project, but will also provide access those already available. The database is web-based and our aim will be to integrate it into the Virtual Atomic and Molecular Data Centre (VAMDC) project (Dubernet et al. 2010). VAMDC data storage is based on the use of XSAMS. XSAMS (XML Schema for Atoms, Molecules and Solids) (Dubernet et al. 2011) is an XML based data storage protocol which has been designed by the International Virtual Observatory Alliance (IVOA) to meet the needs of astronomers who wish to describe or access molecular (and other) data in distributed datasets world-wide.

XSAMS is both flexible and intuitive, making data manipulation and interpretation significantly easier and less error-prone. However the format is very verbose and in its current form it does not seem suitable for storing massive line lists such as individual lists with more than  $10^{10}$  lines which are to be anticipated from the ExoMol project. This will require the development of new protocols and, presumably, adaptation of XSAMS.

Finally, we note that recent test of for models of water spectra in hot Jupiter exoplanet (Tinetti et al. 2012a) suggest that pressure broadening can have a significant influence, particularly at long wavelengths. This means that pressure broadening parameters, particularly those associated with collisions with  $\text{H}_2$  should also be considered for inclusion in the ExoMol database at some future date.

## 4 CONCLUSION

This paper lays out the scope and methodology for a new project, ExoMol, whose aim is to provide comprehensive line lists of molecular transition frequencies and probabilities. The major aim of this project is to provide the necessary data to model atmospheres and interpret spectra for exoplanets and cool stars. However it is recognised that the line lists will have many other applications within astrophysics and beyond. For example it is our practice not to exclude transitions from our lists simply because they are too high in energy to be thermally occupied. This has already led to the identification of new

class of very vibrationally-hot water emissions in comets (Barber et al. 2009). It is to be anticipated that such data will be important for assigning and modelling maser emissions from high lying or hot states. Similarly the database will be available for modelling what may be observable in exoplanet characterisation missions such as the proposed Exoplanet Characterisation Observatory (EChO) (Tinetti et al. 2012b) or FINESSE (Swain 2010) space-borne telescopes.

The ExoMol project will generate very extensive line lists. These will be documented in the present journal and deposited in the linked Strasbourg data repository. The line lists, and other information about the project, will also be made available via the ExoMol website, [www.exomol.com](http://www.exomol.com).

## ACKNOWLEDGEMENTS

This work is supported by ERC Advanced Investigator Project 267219 and a Royal Society Wolfson Research Merit Award. We thank those astronomers who have shared their insight with us in the formulation of this project and in particular Giovanna Tinetti, Peter Bernath, Caitlin Griffiths, Yakiv Pavlenko and Mark Swain. We also thank the present members of the ExoMol team, Ala'a Azzam, Bob Barber, Christian Hill, Lorenzo Lodi, Andrei Patrascu, Oleg Polyansky and Clara Sousa-Silva, as well as Attila Császár and Andrew Kerridge for many helpful discussions on the project, and Olga Yurchenko for her help.

## REFERENCES

- Ahmed K., Balint-Kurti G., Western C., 2004, *J. Chem. Phys.*, 121, 10041
- Albert S., Bauerecker S., Boudon V., Brown L. R., Champion J. P., Loete M., Nikitin A., Quack M., 2009, *Chem. Phys.*, 356, 131
- Allard F., Hauschildt P. H., Alexander D. R., Starrfield S., 1997, *Annu. Rev. Astron. Astrophys.*, 35, 137
- Allard F., Hauschildt P. H., Miller S., Tennyson J., 1994, *ApJ*, 426, L39
- Allard F., Hauschildt P. H., Schwenke D., 2000, *ApJ*, 540, 1005
- Alvarez R., Plez B., 1998, *A&A*, 330, 1109
- Amiot C., Azaroual E. M., Luc P., R V., 1995, *J. Chem. Phys.*, 102, 4375
- Amyay B., Fayt A., Herman M., 2011, *J. Chem. Phys.*, 135, 234305
- Aquilante F. et al., 2010, *J. Comput. Chem.*, 31, 224
- Asvany O., Hugo E., Schlemmer S., Muller F., Kuhnemann F., Schiller S., Tennyson J., 2007, *J. Chem. Phys.*, 127, 154317
- Bailey J., 2009, *Icarus*, 201, 444
- Baker J., Bramble S. K., Hamilton P. A., 1993, *Chem. Phys. Lett.*, 213, 297
- Banerjee D. P. K., Barber R. J., Ashok N. K., Tennyson J., 2005, *ApJ*, 672, L141
- Barber R. J., Banerjee D. P. K., Ashok N. K., Tennyson J., 2007, in *Astron. Soc. Pacific Conf. Series*, Vol. 363, *The Nature of V838 Mon and its Light Echo*, Corradi R. L. M., Munari U., eds., pp. 840–848
- Barber R. J., Harris G. J., Tennyson J., 2002, *J. Chem. Phys.*, 117, 11239
- Barber R. J., Miller S., Dello Russo N., Mumma M. J., Tennyson J., Guio P., 2009, *MNRAS*, 398, 1593
- Barber R. J., Tennyson J., Harris G. J., Tolchenov R. N., 2006, *MNRAS*, 368, 1087
- Barletta P., Császár A. G., Quiney H. M., Tennyson J., 2002, *Chem. Phys. Lett.*, 361, 121
- Bean J. L., Miller-Ricci Kempton E., Homeier D., 2010, *Nature*, 468, 669
- Beaulieu J. P. et al., 2011, *ApJ*, 731, 18
- Béjar V. J. S., Zapatero Osorio M. R., Rebolo R., 1999, *ApJ*, 521, 671
- Bergeron P., Ruiz M., Leggett S., 1997, *ApJS*, 108, 339
- Bernath P. F., Colin R., 2009, *J. Mol. Spectrosc.*, 257, 20
- Berrington K. A., Burke P. G., Butler K., Seaton M. J., Storey P. J., Taylor K. T., Yan Y., 1987, *J. Phys.B: At. Mol. Opt. Phys.*, 20, 6379
- Bornhauser P., Knopp G., Gerber T., Radi P. P., 2010, *J. Mol. Spectrosc.*, 262, 69
- Bornhauser P., Sych Y., Knopp G., Gerber T., Radi P. P., 2011, *J. Chem. Phys.*, 134
- Brandes G. R., Galehouse D. C., 1985, *J. Mol. Spectrosc.*, 109, 345
- Burrows A., Dulick M., Bauschlicher C. W., Bernath P. F., Ram R. S., Sharp C. M., Milsom J. A., 2005, *ApJ*, 624, 988
- Burrows A., Ram R. S., Bernath P., Sharp C. M., Milsom J. A., 2002, *ApJ*, 577, 986
- Bykov A. D., Lavrientieva N. N., Mishina T. P., Sinitsa L. N., Barber R. J., Tolchenov R. N., Tennyson J., 2008, *J. Quant. Spectrosc. Radiat. Transf.*, 109, 1834
- Carter S., Bowman J. M., 1998, *J. Chem. Phys.*, 108, 4397
- Charbonneau D., Brown T., Noyes R., Gilliland R., 2002, *ApJ*, 568, 377
- Chesnokova T. Y., Voronin B. A., Bykov A. D., Zhuravleva T. B., Kozodoev A. V., Lugovskoy A. A., Tennyson J., 2009, *J. Mol. Spectrosc.*, 256, 41

- Coppola C. M., Lodi L., Tennyson J., 2011, MNRAS, 415, 487
- Cours T., Rosmus P., Tyuterev V. G., 2002, J. Chem. Phys., 117, 5192
- Dello Russo N., Bonev B. P., DiSanti M. A., Gibb E. L., Mumma M. J., Magee-Sauer K., Barber R. J., Tennyson J., 2005, ApJ, 621, 537
- Dello Russo N., DiSanti M. A., Magee-Sauer K., Gibb E. L., Mumma M. J., Barber R. J., Tennyson J., 2004, Icarus, 168, 186
- Depraz S., Perrin M. Y., Rivire P., Soufiani A., 2012, J. Quant. Spectrosc. Radiat. Transf., 113, 14
- Des Marais D. J. et al., 2002, Astrobiology, 2, 153
- Dinelli B. M. et al., 1997, Icarus, 126, 107
- Doss N., Tennyson J., Saenz A., Jonsell S., 2006, Phys. Rev. C, 73, 025502
- Dubernet M. L. et al., 2010, J. Quant. Spectrosc. Radiat. Transf., 111, 2151
- Dubernet M. L., Humbert D., Ralchenko Y., Roueff E., Schultz D. R., Chung H.-K., Braams B., 2011, XSAMS: XML Schema for Atomic, Molecular and Solid Data, version 0.1.1. International Atomic Energy Authority, <http://www-amdis.iaea.org/xsams/documents/>
- Dulick M., Bauschlicher C., Burrows A., Sharp C., Ram R., Bernath P., 2003, ApJ, 594, 651
- Engel A. E., Doss N., Harris G. J., Tennyson J., 2005, MNRAS, 357, 471
- Freedman R. S., Marley M. S., Lodders K., 2008, ApJS, 174, 504
- Furtenbacher T., Császár A. G., Tennyson J., 2007, J. Mol. Spectrosc., 245, 115
- Galehouse D. C., Brault J. W., Davis S. P., 1980, ApJS, 42, 241
- Geballe T. R., Kulkarni S. R., Woodward C. E., Sloan G. C., 1996, ApJ, 467, L101
- Glover S. C. O., Savin D. W., 2009, MNRAS, 393, 911
- Goorvitch D., 1994, ApJS, 95, 535
- Gustafsson B., Edvardsson B., Eriksson K., Jørgensen U. G., Nordlund A., Plez B., 2008, A&A, 486, 951
- Gustavsson T., Amiot C., Verges J., 1991, J. Mol. Spectrosc., 145, 56
- Hargreaves R. J., Hinkle K. H., Bauschlicher, Jr. C. W., Wende S., Seifahrt A., Bernath P. F., 2010, AJ, 140, 919
- Harris G. J., Larner F. C., Tennyson J., Kaminsky B. M., Pavlenko Y. V., Jones H. R. A., 2008, MNRAS, 390, 143
- Harris G. J., Lynas-Gray A., Miller S., Tennyson J., 2004, ApJ, 617, L143
- Harris G. J., Pavlenko Y. V., Jones H. R. A., Tennyson J., 2003, MNRAS, 344, 1107
- Harris G. J., Polyansky O. L., Tennyson J., 2002, ApJ, 578, 657
- Harris G. J., Tennyson J., Kaminsky B. M., Pavlenko Y. V., Jones H. R. A., 2006, MNRAS, 367, 400
- Hefferlin R., Kuznetsova L. A., 1999, J. Quant. Spectrosc. Radiat. Transf., 62, 765
- Hill C., Yurchenko S. N., Tennyson J., ????, Icarus
- Hill J. G., Mazumder S., Peterson K. A., 2010, J. Chem. Phys., 132
- Hocking W. H., Gerry M. C. L., Merer A. J., 1979, Can. J. Phys., 57, 54
- Homeier D., Hauschildt P., Allard F., 2003, in Astronomical Society of the Pacific Conference Series, Vol. 288, Stellar atmosphere modeling, Hubeny I., Mihalas D., Werner K., eds., pp. 357–360
- Hougen J. T., Bunker P. R., Johns J. W. C., 1970, J. Mol. Spectrosc., 34, 136
- Huang X., Schwenke D. W., Tashkun S. A., Lee T. J., 2012, J. Chem. Phys.
- Irwin A., 1981, ApJS, 45, 621
- Jacquinet-Husson N. et al., 2011, J. Quant. Spectrosc. Radiat. Transf., 112, 2395
- Joester J. A., Nakajima M., Reilly N. J., Kokkin D. L., Nauta K., Kable S. H., Schmidt T. W., 2007, J. Chem. Phys., 127, 214303
- Jones H. R. A., Pavlenko Y., Viti S., Tennyson J., 2002, MNRAS, 330, 675
- Jørgensen U., Almlöf J., Gustafsson B., Larsson M., Siegbahn P., 1985, J. Chem. Phys., 83, 3034
- Jørgensen U. G., 1994, A&A, 284, 179
- Jørgensen U. G., 1996, in IAU Symposium, Vol. 178, Molecules in Astrophysics: Probes & Processes, E. F. van Dishoeck, ed., p. 441
- Jørgensen U. G., Almlöf J., Siegbahn P. E. M., 1989, ApJ, 343, 554
- Jørgensen U. G., Jensen P., Sørensen G. O., Aringer B., 2001, A&A, 372, 249
- Kaledin L. A., McCord J. E., Heaven M. C., 1995, J. Mol. Spectrosc., 173, 499
- Kaltenegger L. et al., 2010, Astrobiology, 10, 89
- Kokkin D. L., Bacskay G. B., Schmidt T. W., 2007, J. Chem. Phys., 126, 084302
- Kokkin D. L., Reilly N. J., Morris C. W., Nakajima M., Nauta K., Kable S. H., Schmidt T. W., 2006, J. Chem. Phys., 125, 231101
- Koskinen T. T., Aylward A. D., Miller S., 2007, Nature, 450, 845
- Kozin I. N., Law M. M., Tennyson J., Hutson J. M., 2004, Comput. Phys. Commun., 163, 117
- Kozin I. N., Law M. M., Tennyson J., Hutson J. M., 2005, J. Chem. Phys., 122, 064309

- Kranendonk L. A. et al., 2007, *Opt. Express*, 15, 15115
- Krechel H. et al., 2002, *Phys. Rev. A*, 66, 052509
- Kreckel H., Schwalm D., Tennyson J., Wolf A., Zajfman D., 2004, *New J. Phys.*, 6, 151.1
- Kurucz R. L., 2011, *Can. J. Phys.*, 89, 417
- Langhoff S. R., Bauschlicher C. W., 1993, *Chem. Phys. Lett.*, 211, 305
- Launila O., Berg L.-E., 2011, *J. Mol. Spectrosc.*, 265, 10
- Le Roy R. J., 2006, DPotFit 1.1 A Computer Program for Fitting Diatomic Molecule Spectral Data to Potential Energy Functions. University of Waterloo Chemical Physics Research Report CP-662R, <http://leroy.uwaterloo.ca/programs/>
- Le Roy R. J., 2007, LEVEL 8.0 A Computer Program for Solving the Radial Schrödinger Equation for Bound and Quasibound Levels. University of Waterloo Chemical Physics Research Report CP-663, <http://leroy.uwaterloo.ca/programs/>
- Le Sueur C. R., Miller S., Tennyson J., Sutcliffe B. T., 1992, *Mol. Phys.*, 76, 1147
- Lindermeir E., Beier K., 2012, *J. Quant. Spectrosc. Radiat. Transf.*, 113, 1575
- Linton C., 1974, *J. Mol. Spectrosc.*, 50, 235
- Lischka H. et al., 2001, *Phys. Chem. Chem. Phys.*, 3, 664
- Lodi L., Tennyson J., 2012, *J. Quant. Spectrosc. Radiat. Transf.*, 113, 850
- Lodi L., Tennyson J., Polyansky O. L., 2011, *J. Chem. Phys.*, 135, 034113
- Lodi L. et al., 2008, *J. Chem. Phys.*, 128, 044304
- Lorono Gonzalez M. A. et al., 2010, *J. Quant. Spectrosc. Radiat. Transf.*, 111, 2265
- Lynas-Gray A. E., Miller S., Tennyson J., 1995, *J. Mol. Spectrosc.*, 169, 458
- Lyubchik Y., Jones H. R. A., Pavlenko Y. V., Martin E., McLean I., Prato L., Barber R. J., Tennyson J., 2007, *A&A*, 473, 257
- Ma L., Cai W., Caswell A. W., Kraetschmer T., Sanders S. T., Roy S., Gord J. R., 2009, *Opt. Express*, 17, 8602
- Marian C. M., 1995, *Ber. Bunsen. Phys. Chem.*, 99, 254
- Marian C. M., 2001, in *Rev. Comput. Chem.*, Lipkowitz K., Boyd D., eds., Vol. 17, John Wiley & Sons, Inc., pp. 99–204
- Mathar R. J., 2007, *J. Optics A: Pure Appl. Optics*, 9, 470
- Matsuura M. et al., 2005, *A&A*, 434, 691
- Mátyus E., Czakó G., Sutcliffe B. T., Császár A. G., 2007, *J. Chem. Phys.*, 127, 084102
- Mellau G. C., 2011a, *J. Chem. Phys.*, 134
- Mellau G. C., 2011b, *J. Mol. Spectrosc.*, 269, 77
- Miller S. et al., 2000, *Phil. Trans. Royal Soc. London A*, 358, 2485
- Miller S., Stallard T., Melin H., Tennyson J., 2010, *Faraday Discuss.*, 147, 283
- Moudens A., Georges R., Benidar A., Amyay B., Herman M., Fayt A., Plez B., 2011, *J. Quant. Spectrosc. Radiat. Transf.*, 112, 540
- Müller H. S. P., Schlöder F., Stutzki J., Winnewisser G., 2005, *J. Mol. Struct.*, 742, 215
- Na C. Y., Esposito L. W., Skinner T. E., 1990, *J. Geophys. Res. A*, 95, 7485
- Nakajima M., Joester J. A., Page N. I., Reilly N. J., Bacsikay G. B., Schmidt T. W., Kable S. H., 2009, *J. Chem. Phys.*, 131, 044301
- Neale L., Miller S., Tennyson J., 1996, *ApJ*, 464, 516
- Neale L., Tennyson J., 1995, *ApJ*, 454, L169
- Northrup F., Sears T., 1990, *J. Opt. Soc. Am. B*, 7, 1924
- Oka T., Epp E., 2004, *ApJ*, 613, 349
- Ovsyannikov R. I., Thiel W., Yurchenko S. N., Carvajal M., Jensen P., 2008, *J. Mol. Spectrosc.*, 252, 121
- Partridge H., Schwenke D. W., 1997, *J. Chem. Phys.*, 106, 4618
- Pavlenko Y. V., Harris G. J., Tennyson J., Jones H. R. A., Brown J. M., Hill C., Yakovina L. A., 2008, *MNRAS*, 386, 1338
- Pickett H. M., Poynter R. L., Cohen E. A., Delitsky M. L., Pearson J. C., Müller H. S. P., 1998, *J. Quant. Spectrosc. Radiat. Transf.*, 60, 883
- Plez B., 1998, *A&A*, 337, 495
- Polyansky O. L., Császár A. G., Shirin S. V., Zobov N. F., Barletta P., Tennyson J., Schwenke D. W., Knowles P. J., 2003, *Science*, 299, 539
- Polyansky O. L., Zobov N. F., Viti S., Tennyson J., Bernath P. F., Wallace L., 1997, *ApJ*, 489, L205
- Ram R. S., Bernath P. F., 2011, *ApJS*, 194
- Ram R. S., Bernath P. F., Wallace L., 1996, *ApJS*, 107, 443
- Ram R. S., Wallace L., Hinkle K., Bernath P. F., 2010, *ApJS*, 188, 500
- Rohlfing E., 1989, *J. Chem. Phys.*, 91, 4531
- Rohlfing E., Goldsmith J., 1989, *J. Chem. Phys.*, 90, 6804
- Rohlfing E., Goldsmith J., 1990, *J. Opt. Soc. Am. B*, 7, 1915
- Ross A., Crozet P., Richard C., Harker H., Ashworth S., 2012, *Mol. Phys.*, *in press*, XXX, XXX



- Rothman L. S. et al., 2009, *J. Quant. Spectrosc. Rad. Transf.*, 110, 533
- Rothman L. S. et al., 2010, *J. Quant. Spectrosc. Radiat. Transf.*, 111, 2139
- Rothman L. S., Wattson R. B., Gamache R. R., Schroeder J., McCann A., 1995, *Proc. Soc. Photo. Opt. Instrum. Eng.*, 2471, 105
- Saha S., Western C. M., 2006, *J. Chem. Phys.*, 125, 224307
- Sauval A. J., Tatum J. B., 1984, *ApJS*, 56, 193
- Schleicher D. R. G., Galli D., Palla F., Camenzind M., Klessen R. S., Bartelmann M., Glover S. C. O., 2008, *A&A*, 490, 521
- Schmidt T. W., Bacskay G. B., 2007, *J. Chem. Phys.*, 127, 234310
- Schmidt T. W., Bacskay G. B., 2011, *J. Chem. Phys.*, 134, 224311
- Schwenke D., 1998, *Faraday Discuss.*, 109, 321
- Schwenke D., 2002, *Spectra Chimica Acta A*, 58, 849
- Schwenke D. W., Partridge H., 2000, *J. Chem. Phys.*, 113, 6592
- Seaton M., 1987, *J. Phys.B: At. Mol. Opt. Phys.*, 20, 6363
- Seaton M. J., 2005, *MNRAS*, 362, L1
- Sharp C. M., Burrows A., 2007, *Astrophys. J. Suppl. Ser.*, 168, 140
- Sharpe S. W., Johnson T. J., Sams R. L., Chu P. M., Rhoderick G. C., Johnson P. A., 2004, *Appl. Spectrosc.*, 58, 1452
- Simard B., Hackett P. A., 1991, *J. Mol. Spectrosc.*, 148, 128
- Simeckova M., Jacquemart D., Rothman L., Gamache R., Goldman A., 2006, *J. Quant. Spectrosc. Radiat. Transf.*, 98, 130
- Skory S., Weck P. F., Stancil P. C., Kirby K., 2003, *ApJS*, 148, 599
- Sochi T., Tennyson J., 2010, *MNRAS*, 405, 2345
- Sousa-Silva C., Yurchenko S. N., Tennyson J., 2012, *J. Mol. Spectrosc.*
- Spirko V., Mengel M., Jensen P., 1997, *J. Mol. Spectrosc.*, 183, 129
- Stancil P. C., 1994, *ApJ*, 430, 360
- Stevenson K. B. et al., 2010, *Nature*, 464, 1161
- Swain M. R., 2010, in *Bulletin of the American Astronomical Society*, Vol. 42, AAS/Division for Planetary Sciences Meeting Abstracts, p. 1064
- Swain M. R., Vasisht G., Tinetti G., 2008, *Nature*, 452, 329
- Tanabashi A., Hirao T., Amano T., Bernath P. F., 2007, *ApJS*, 169, 472
- Tashkun S. A., Perevalov V. I., 2011, *J. Quant. Spectrosc. Radiat. Transf.*, 112, 1403
- Tennyson J., 2011, *WIREs Comput. Mol. Sci.*, 10.1002/wcms.94
- Tennyson J. et al., 2010, *J. Quant. Spectrosc. Radiat. Transf.*, 111, 2160
- Tennyson J. et al., 2009, *J. Quant. Spectrosc. Radiat. Transf.*, 110, 573
- Tennyson J., Harris G. J., Barber R. J., La Delfa S., Voronin B. A., Kaminsky B. M., Pavlenko Y. V., 2007, *Mol. Phys.*, 105, 701
- Tennyson J., Kostin M. A., Barletta P., Harris G. J., Polyansky O. L., Ramanlal J., Zobov N. F., 2004, *Comput. Phys. Commun.*, 163, 85
- Tennyson J., Sutcliffe B. T., 1982, *J. Chem. Phys.*, 77, 4061
- Tennyson J., Yurchenko S. N., 2012, *MNRAS*
- The Opacity Project Team, Berrington K. A., 1994, *The Opacity Project*, volume 2. Insitute of Physics Publishing
- The Opacity Project Team, Seaton M., 1995, *The Opacity Project*, volume 1. Insitute of Physics Publishing
- Tinetti G. et al., 2012a, *Experimental Astronomy*
- Tinetti G., Tennyson J., Griffiths C. A., Waldmann I., 2012b, *Phil. Trans. Royal Soc. London A*, 370, 2749
- Tinetti G. et al., 2007, *Nature*, 448, 169
- Tyuterev V. G., Tashkun S. A., Schwenke D. W., 2001, *Chem. Phys. Lett.*, 348, 223
- Urru A., Kozin I. N., Mulas G., Braams B. J., Tennyson J., 2010, *Mol. Phys.*, 108, 1973
- Uttenthaler S., Aringer B., Lebzelter T., Kaeuffl H. U., Siebenmorgen R., Smette A., 2008, *ApJ*, 682, 509
- Vallon R., Richard C., Crozet P., Wannous G., Ross A., 2009, *ApJ*, 696, 172
- Vidler M., Tennyson J., 2000, *J. Chem. Phys.*, 113, 9766
- Viti S., Tennyson J., Polyansky O. L., 1997, *MNRAS*, 287, 79
- Voronin B. A., Tennyson J., Tolchenov R. N., Lugovskoy A. A., Yurchenko S. N., 2010, *MNRAS*, 402, 492
- Wang L., Mondelain D., Kassi S., Campargue A., 2012, *J. Quant. Spectrosc. Radiat. Transf.*, 113, 47
- Wang Y., Braams B. J., Bowman J. M., Carter S., Tew D. P., 2008, *J. Chem. Phys.*, 128
- Warmbier R., Schneider R., Sharma A. R., Braams B. J., Bowman J. M., Hauschildt P. H., 2009, *A&A*, 495, 655
- Watson J. K. G., 1968, *Molec. Phys.*, 15, 479
- Wattson R. B., Rothman L. S., 1992, *J. Quant. Spec. Radiative. Transf.*, 48, 763
- Weck P. F., Schweitzer A., Kirby K., Hauschildt P. H., Stancil P. C., 2004, *ApJ*, 613, 567
- Weck P. F., Schweitzer A., Stancil P. C., Hauschildt P. H., Kirby K., 2003a, *ApJ*, 584, 459

- Weck P. F., Schweitzer A., Stancil P. C., Hauschildt P. H., Kirby K., 2003b, *ApJ*, 582, 1059
- Weck P. F., Stancil P. C., Kirby K., 2003c, *ApJ*, 582, 1263
- Weck P. F., Stancil P. C., Kirby K., 2003d, *J. Chem. Phys.*, 118, 9997
- Wende S., Reiners A., Seifahrt A., Bernath P. F., 2010, *A&A*, 523
- Werner H. J., Knowles P. J., Lindh R., Manby F. R., Schütz M., et al., 2010, *MOLPRO*, a package of ab initio programs.  
See <http://www.molpro.net/>
- Yachmenev A., Yurchenko S. N., Jensen P., Thiel W., 2011, *J. Chem. Phys.*, 134
- Yurchenko S. N., Barber R. J., Tennyson J., 2011a, *MNRAS*, 413, 1828
- Yurchenko S. N., Barber R. J., Tennyson J., Thiel W., Jensen P., 2011b, *J. Mol. Spectrosc.*, 268, 123
- Yurchenko S. N., Barber R. J., Yachmenev A., Thiel W., Jensen P., Tennyson J., 2009, *J. Phys. Chem. A*, 113, 11845
- Yurchenko S. N., Carvajal M., Jensen P., Herregodts F., Huet T. R., 2003, *Chem. Phys.*, 290, 59
- Yurchenko S. N., Carvajal M., Thiel W., Jensen P., 2006, *J. Mol. Spectrosc.*, 239, 71
- Yurchenko S. N., Thiel W., Jensen P., 2007, *J. Mol. Spectrosc.*, 245, 126
- Yurchenko S. N., Voronin B. A., Tolchenov R. N., Doss N., Naumenko O. V., Thiel W., Tennyson J., 2008, *J. Chem. Phys.*, 128, 044312
- Zaharova J., Tamanis M., Ferber R., Drozdova A. N., Pazyuk E. A., Stolyarov A. V., 2009, *Phys. Rev. A*, 79, 012508